**Baleen Analytic Tool: a noodling framework for the generalized search and analysis of heterogeneous data sources**

**Summary:**

Many tools allow for the auditing and analysis of standardized data. Data Warehouses (DW) and Data Marts are constructed, by consultants, at great expense to the customer – these data archives are created for the purpose of supporting business intelligence and in some cases auditing.

What the **Baleen Analytic Tool (BAT)** provides is a simple, general, efficient and fast way of analyzing, archiving and performing complex analytics on data which is both structured and unstructured. It achieves this by reducing the non-random information space (what we call Data Atomization) to its primary basic forms – sequences and trees.

We also recognize that the ATOMIC value of any data element can be reduced to one of the following:

1. Strings or Text (also known as sequential data).
2. Numbers (real and natural)
3. Temporal (date/time)
4. File Reference (image, audio, film, etc.)

However, in the world of relational data modeling and Kimball/Inman DW methodologies, this kind of reductionism is frowned upon. Why? Because, frankly, the first class citizens of information science are the atomic values which comprise business objects and non-structured sequential data (text, music, film, image).

Trees or records are the basic form of structured data most often found in a contemporary data warehouse. Sequences (like free form text) comprise the other part of data found in a active data warehouse, OLTP or data mart.

BAT reduces both Trees and Sequences down to a form which is roughly equivalent, and both types of structures share the same common value space dictionary.